Estatística Básica II

PROBABILIDADE

TUANY CASTRO

Variáveis aleatórias

- Características de interesse em uma população;
- **Exemplo 1:** em uma linha de usinagem de peças estamos interessados em controlar o diâmetro das peças produzidas. Neste caso, o resultado da medição do diâmetro é a variável aleatória de interesse;
- **Exemplo 2:** em um ensaio clínico, estamos interessados em avaliar o tempo de vida dos pacientes e neste caso, a tempo de vida corresponde à variável aleatória.

Função de probabilidade

X é uma variável aleatória discreta se assume valores inteiros.

Exemplo: Um centro de saúde constatou que, para as famílias de uma região, 20% não têm filhos, 30% têm um filho, 35% têm dois filhos, 5% têm 3 filhos, 5% têm 4 filhos e 5% tem 5 filhos.

Seja X o número de filhos de uma família → X é uma **variável aleatória discreta**

Função de probabilidade de X:

X	**0**	**1**	**2**	**3**	**4**	**5**
P(X=x)	0,20	0,30	0,35	0,05	0,05	0,05

Função acumulada de probabilidade

A função de distribuição ou função acumulada de probabilidade de uma variável aleatória discreta X é definida como:

$$F(x) = P(X \leq x)$$

Exemplo: X é o número de filhos de uma família e tem a seguinte função de probabilidade:

X	**0**	**1**	**2**	**3**	**4**	**5**
P(X=x)	0,20	0,30	0,35	0,05	0,05	0,05

Exemplo

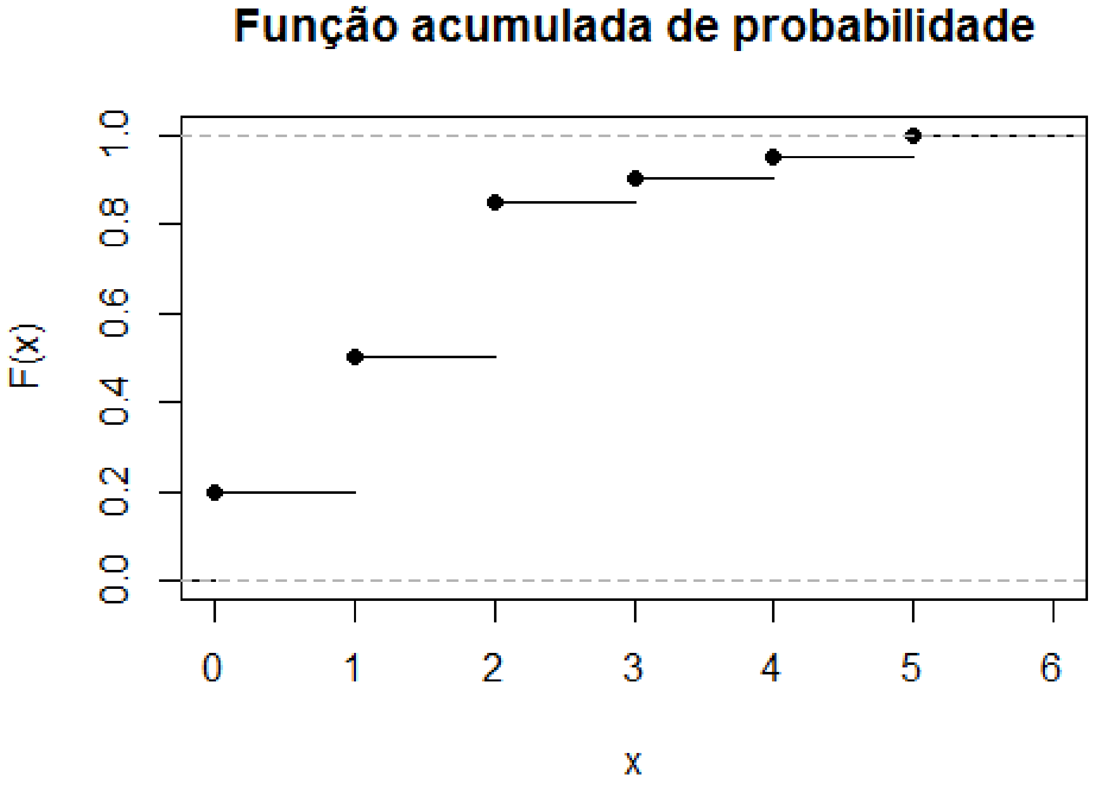
Qual a função acumulada de probabilidade de X?

- $F(0) = P(X \leq 0) = P(X = 0) = 0{,}20$
- $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 0{,}20 + 0{,}30 = 0{,}50$
- $F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}20 + 0{,}30 + 0{,}35 = 0{,}85$
- $F(3) = P(X \leq 3) = P(X \leq 2) + P(X = 3) = 0{,}85 + 0{,}05 = 0{,}90$
- $F(4) = P(X \leq 4) = P(X \leq 3) + P(X = 4) = 0{,}90 + 0{,}05 = 0{,}95$
- $F(5) = P(X \leq 5) = P(X \leq 4) + P(X = 5) = 0{,}95 + 0{,}05 = 1{,}0$

Exemplo

Função acumulada de probabilidade de X:

$$F(x) = \begin{cases} 0, se\ x < 0 \\ 0{,}20\ se\ 0 \leq x < 1 \\ 0{,}50\ se\ 1 \leq x < 2 \\ 0{,}85\ se\ 2 \leq x < 3 \\ 0{,}90\ se\ 3 \leq x < 4 \\ 0{,}95\ se\ 4 \leq x < 5 \\ 1\ se\ x \geq 5 \end{cases}$$



Esperança

A média, valor esperado ou esperança de uma variável X é dada pela expressão:

$$E(X) = \mu = \sum_{i=1}^{n} x_i * P(X = x_i)$$

Exemplo: X é o número de filhos de uma família e tem a seguinte função de probabilidade:

X	**0**	**1**	**2**	**3**	**4**	**5**
P(X=x)	0,20	0,30	0,35	0,05	0,05	0,05

A **esperança** de X:

$$E(X) = 0*0,20 + 1*0,30 + 2*0,35 + 3*0,05 + 4*0,05 + 5*0,05 = 1,6$$

Variância

A variância de uma variável X é dada pela expressão:

$$Var(X) = \sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2 * P(X = x_i)$$

Exemplo: X é o número de filhos de uma família e tem a seguinte função de probabilidade:

X	**0**	**1**	**2**	**3**	**4**	**5**
P(X=x)	0,20	0,30	0,35	0,05	0,05	0,05

A **variância** de X:

$$\text{Var}(X) = \sigma^2 = (0-1{,}6)^2*0{,}20 + (1-1{,}6)^2*0{,}30 + (2-1{,}6)^2*0{,}35 + (3-1{,}6)^2*0{,}05 + (4-1{,}6)^2*0{,}05 + (5-1{,}6)^2*0{,}05$$

$$\text{Var}(X) = \sigma^2 = 1{,}64$$

Exemplo

Considere o experimento de lançar uma moeda honesta e observar se ocorre cara (C) ou coroa (R). Descreva as funções de probabilidade e probabilidade acumulada da variável número de caras em dois lançamentos dessa moeda. Qual a média e a variância dessa variável?

Espaço amostral: $\Omega = \{CC, CR, RC, RR\}$

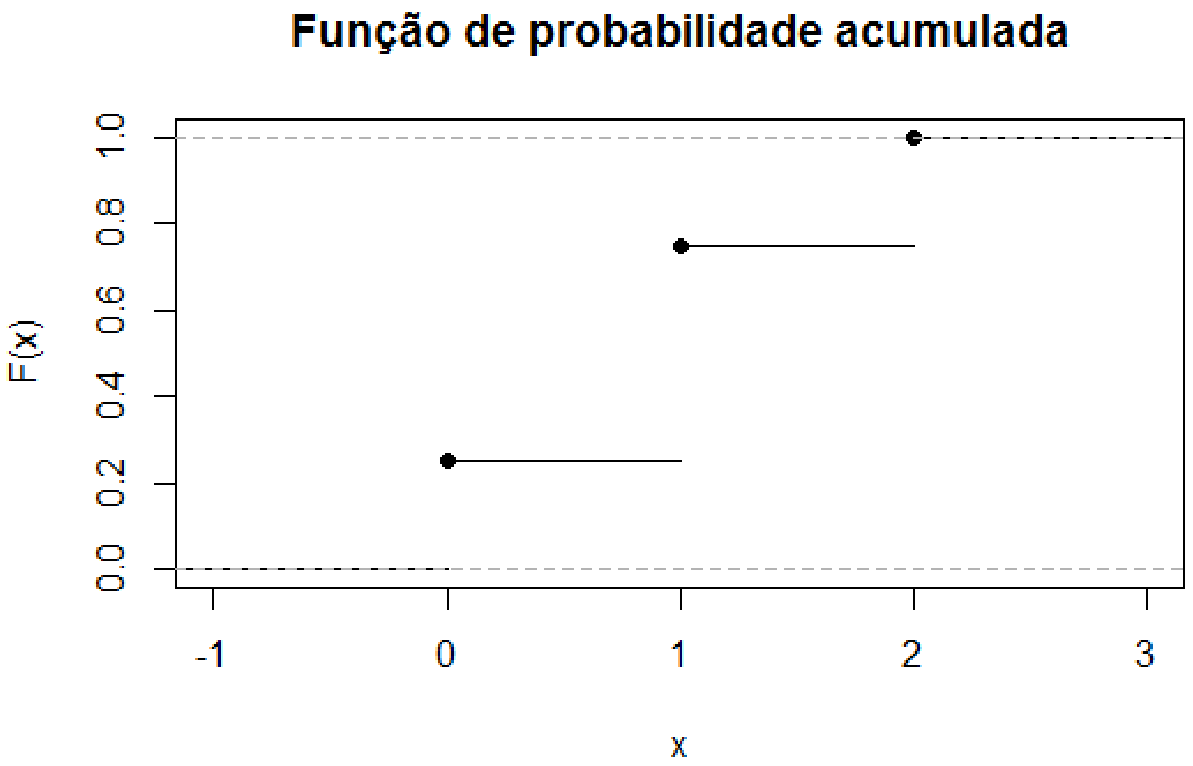
X: número de caras nos dois lançamentos da moeda (variável aleatória discreta)

$$X = \begin{cases} 0, se\ CC \\ 1, se\ CR\ ou\ RC \\ 2, se\ RR \end{cases}$$

X	**0**	**1**	**2**
P(X=x)	0,25	0,50	0,25

Exemplo

$$F(x) = \begin{cases} 0 \; se \; x < 0 \\ 0{,}25 \; se \; 0 \; \leq x < 1 \\ 0{,}75 \; se \; 1 \; \leq x < 2 \\ 1 \; se \; x \; \geq 2 \end{cases}$$



Exemplo

Média:

$$\mu = 0{,}25 * 0 + 0{,}50 * 1 + 0{,}25 * 2$$

$$\mu = 1$$

Variância:

$$\sigma^2 = 0{,}25 * (0 - 1)^2 + 0{,}50 * (1 - 1)^2 + 0{,}25 * (2 - 1)^2$$

$$\sigma^2 = 0{,}50$$

Exercícios

Num teste de digitação, o tempo em minutos (T) que os candidatos levam para digitar um texto é modelado, de forma aproximada, pela seguinte função de probabilidade:

T	3	4	5	6	7	8	9
P(T=t)	0,1	0,1	0,2	0,2	0,2	0,1	0,1

Determine o valor esperado e a variância do tempo de digitação.

Desafio: O candidato recebe 4 pontos se terminar a digitação em 9 minutos, 5 se terminar em 8 minutos e assim por diante. Determine a média e a variância do número de pontos obtidos no teste.

Modelo Uniforme Discreto

Seja X uma variável aleatória cujos possíveis valores são representados por $x_1, x_2, \dots, x_k$. Dizemos que X segue o modelo Uniforme Discreto se atribui a mesma probabilidade $1/k$ a cada um desses k valores, isto é, sua função de probabilidade é dada por:

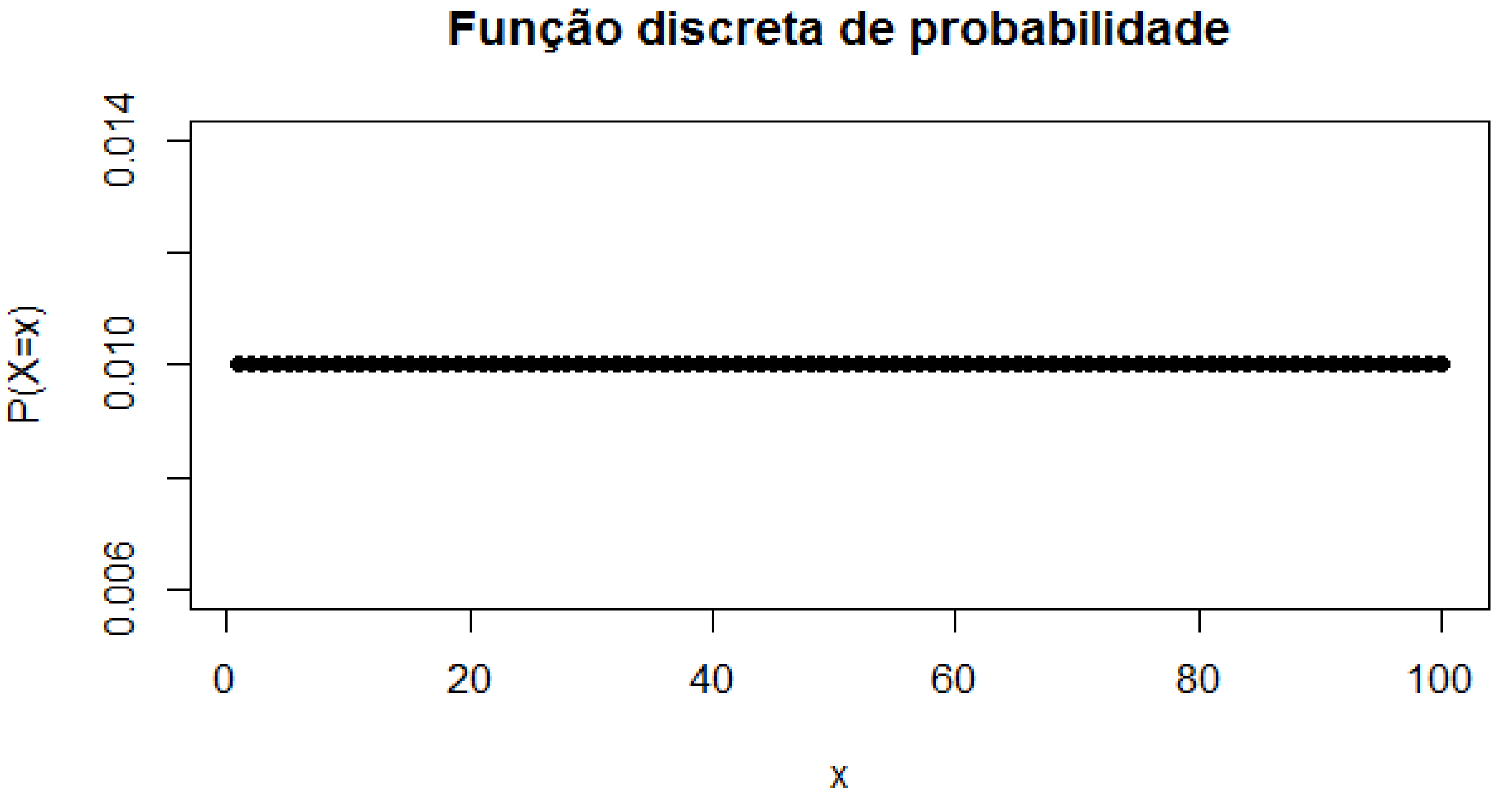
$$P(X = x_i) = \frac{1}{k}$$

Exemplo: Uma rifa tem 100 bilhetes numerados de 1 a 100. Tenho um bilhete de número 21 e meu colega tem outro bilhete de número 68. Quem tem a maior probabilidade de ser sorteado?

Seja X o número sorteado. Os possíveis valores de X vão de 1 a 100.
Assumindo a honestidade da rifa, todos os números tem a mesma probabilidade de ocorrência:

$$P(X = x) = 1/100.$$

Modelo Uniforme Discreto



Modelo Uniforme Discreto

Esperança

$$\boldsymbol{E(X) = \frac{1 + k}{2}}$$

Variância

$$\textbf{Var}(\boldsymbol{X}) = \boldsymbol{\frac{k^2 - 1}{12}}$$

Modelo Uniforme Discreto

Exemplo:

$$\boldsymbol{E(X) = \frac{1+k}{2} = \frac{1+100}{2} = 50,5}$$

$$\textbf{Var}(\boldsymbol{X}) = \boldsymbol{\frac{k^2-1}{12} = \frac{100^2-1}{12} = 833,25}$$

Modelo Bernoulli

Em muitas situações práticas a variável de interesse assume somente dois valores. Essas situações podem ser representadas por respostas do tipo *sucesso-fracasso,* sendo a atribuição de *sucesso* feita de maneira arbitrária porém clara, para evitar ambiguidades.

Dizemos que X segue o modelo Bernoulli se atribui 0 ou 1 à ocorrência de fracasso ou sucesso, respectivamente. Com p representando a probabilidade de *sucesso*, $0 \leq p \leq 1$, sua função discreta de probabilidade é dada por:

x	0	1
P(X=x)	1-p	p

Ou, de modo resumido, $P(X = 0) = 1 - p$ e $P(X = 1) = p$

Modelo Bernoulli

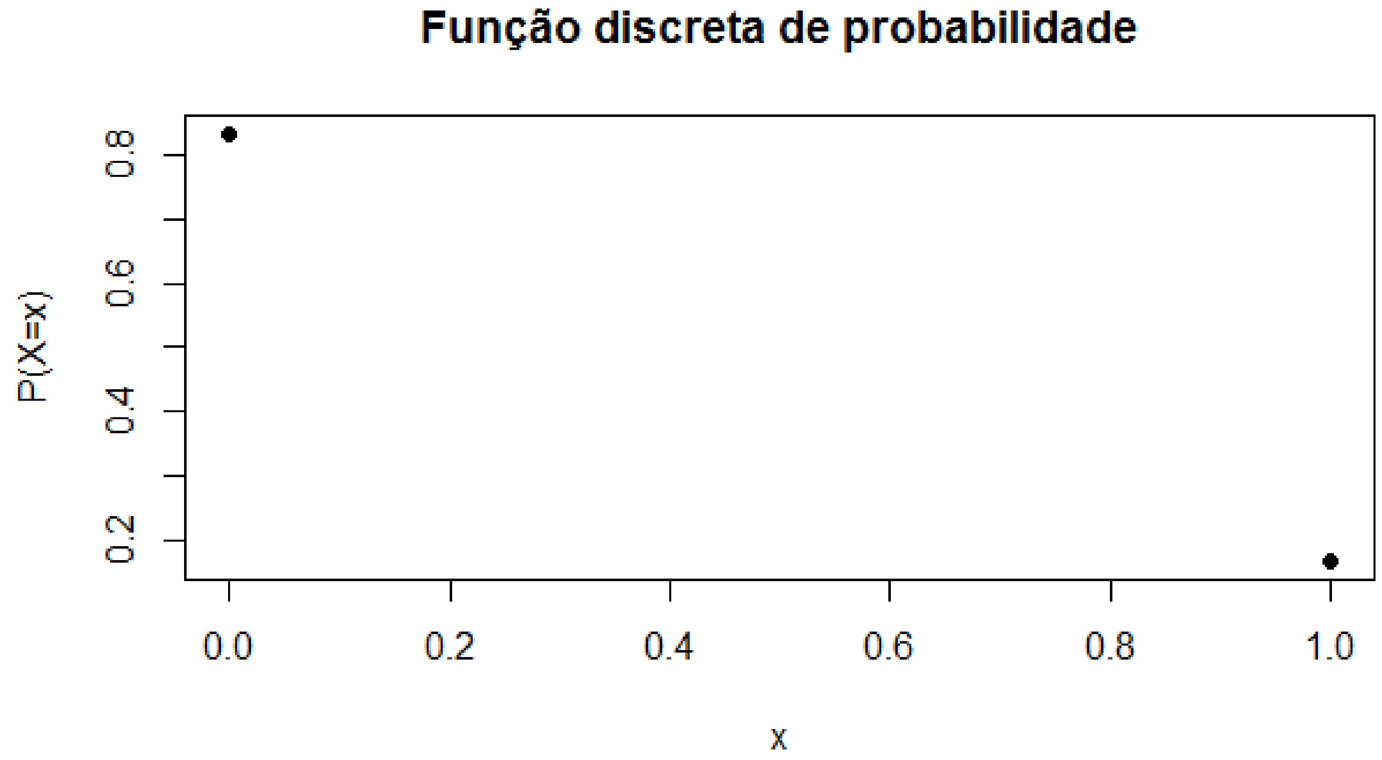
Exemplo: No lançamento de um dado, ocorre a face ou não. Considerando X a variável aleatória que indica a ocorrência de face 5, qual a distribuição de X?

Seja ocorrência de face 5 o sucesso, temos:

$$P(X = 1) = {}^{1}/_{6}$$

$$P(X = 0) = 1 - {}^{1}/_{6} = {}^{5}/_{6}$$

Modelo Bernoulli



Modelo Bernoulli

Esperança

$$\boldsymbol{E(X) = p}$$

Variância

$$\mathbf{Var}(\boldsymbol{X}) = \boldsymbol{p(1-p)}$$

Modelo Bernoulli

Exemplo:

$$E(X) = p = {}^{1}/_{6}$$

$$\text{Var}(X) = p(1 - p) = {}^{1}/_{6}\,(1 - {}^{1}/_{6}) = {}^{5}/_{36}$$

Modelo Binomial

Considere a repetição de n ensaios de Bernoulli independentes e todos com a mesma probabilidade de sucesso p. A variável aleatória que conta o número total de sucessos é denominada Binomial com parâmetros n e p e sua função de probabilidade:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \mathrm{k} = 0, 1, 2, \dots, \mathrm{n}$$

Usamos a notação $X \sim b(n, p)$ para indicar que X segue distribuição binomial com parâmetros n e p.

Modelo Binomial

Exemplo: Suponha que numa linha de produção a probabilidade de se obter uma peça defeituosa (sucesso) é $p = 0{,}1$. Toma-se uma amostra de 10 peças para serem inspecionadas. Qual a probabilidade de se obter:

(A) Uma peça defeituosa?

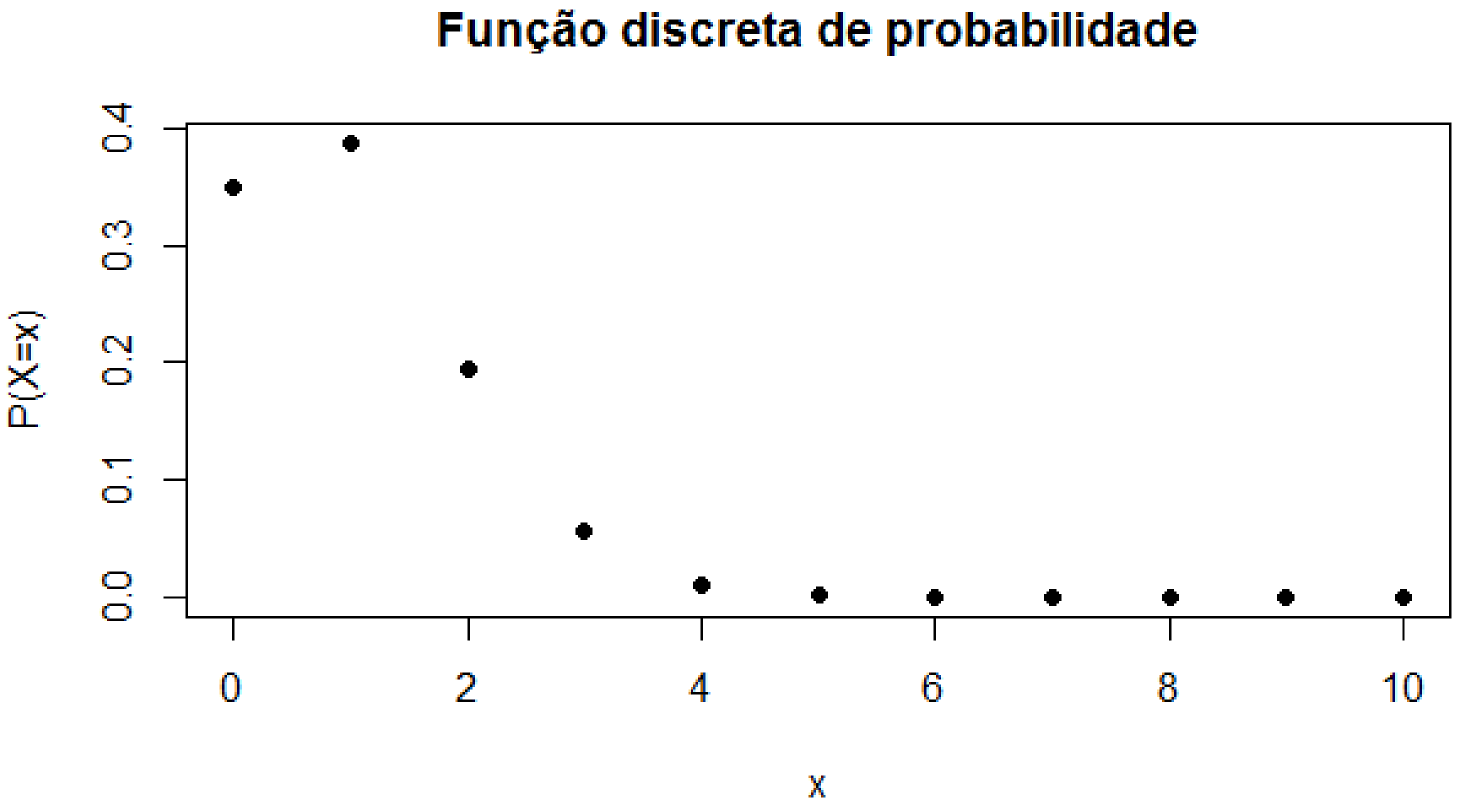
(B) Nenhuma peça defeituosa?

(C) Duas peças defeituosas?

(D) No mínimo duas peças defeituosas?

(E) No máximo duas peças defeituosas?

Modelo Binomial



Modelo Binomial

Esperança

$$\boldsymbol{E(X) = np}$$

Variância

$$\textbf{Var}(\boldsymbol{X}) = \boldsymbol{np(1-p)}$$

Modelo Binomial

Exemplo:

$$\boldsymbol{E(X) = np = 10 * 0,1 = 1}$$

$$\textbf{Var}\boldsymbol{(X) = np(1-p) = 10 * 0,1 * (1-0,1) = 0,9}$$

Exercícios

Exercício 1: Um usuário de transporte coletivo chega pontualmente às 8 horas para pegar o seu ônibus. Devido ao trânsito caótico, a demora pode ser qualquer tempo entre 1 e 20 minutos (admita que o relógio “pule” de minuto em minuto). Pergunta-se:

A) Qual a probabilidade de demorar mais de 10 minutos?

B) Qual a probabilidade de demorar pelo menos 5 mas não mais de 10 minutos?

C) Qual a probabilidade da demora não chegar a 5 minutos?

Exercício 2: Uma certa doença pode ser curada através de procedimento cirúrgico em 80% dos casos. Dentre os que têm essa doença, sorteamos 15 pacientes que serão submetidos à cirurgia. Responda qual é a probabilidade de:

A) Todos serem curados?

B) Pelo menos 10 serem curados?

Exercício

Exercício 3: Um time paulista de futebol tem probabilidade 0,92 de vitória sempre que joga. Se o time atuar 4 vezes, determine a probabilidade de que vença:

A) Todas as partidas.

B) Exatamente 2 partidas.

C) Pelo menos uma partida.

D) No máximo 3 partidas.

E) Mais da metade das partidas.

Exercício 4: Uma vacina contra gripe é eficiente em 70% dos casos. Sorteamos, ao acaso, 20 dos pacientes vacinados e pergunta-se a probabilidade de obter:

A) Pelo menos 18 imunizados.

B) No máximo 4 imunizados.

Modelo Poisson

O modelo Poisson tem sido muito utilizado em experimentos físicos e biológicos, em que a variável aleatória conta o número de ocorrências de um evento numa determinada unidade de tempo ou espaço. Chamamos λ o parâmetro que representa a frequência média ou esperada de ocorrências, a função de probabilidade dessa distribuição é dada por:

$$P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}, \mathrm{k} = 0, 1, 2, \dots$$

Usamos a notação $X \sim Po(\lambda)$ para indicar que X segue distribuição poisson com parâmetro λ.

Modelo Poisson

Exemplo: Suponha que o número de partículas radioativas alfa emitidas por minuto seja uma variável aleatória seguindo o modelo Poisson com parâmetro 5, isto é, a taxa média de ocorrência é de 5 emissões a cada minuto. Calculemos a probabilidade de:

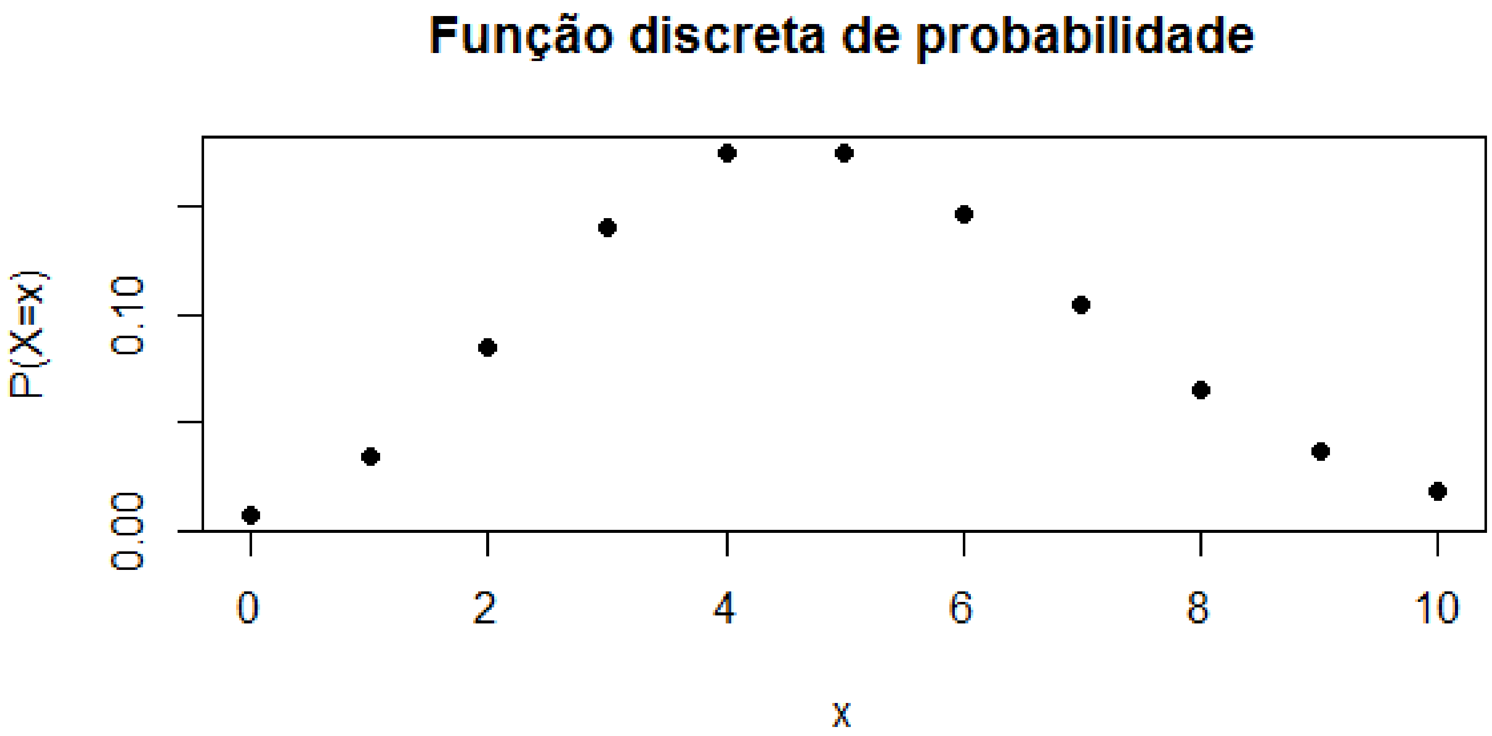
A) Ocorrer exatamente 1 emissão em um minuto.

B) Não ocorrer emissão em um minuto.

C) Ocorrerem duas ou mais emissão em um minuto.

D) Ocorrerem no máximo 3 emissões em um minuto.

Modelo Poisson



Modelo Poisson

Esperança

$$E(X) = \lambda$$

Variância

$$\text{Var}(X) = \lambda$$

Modelo Poisson

Exemplo:

$$\boldsymbol{E(X) = \lambda = 5}$$

$$\textbf{Var}(\boldsymbol{X}) = \boldsymbol{\lambda} = \mathbf{5}$$

Exercícios

Exercício 5: A aplicação de fundo anti-corrosivo em chapas de aço de 1 m^2 é feita mecanicamente e pode produzir defeitos (pequenas bolhas na pintura), de acordo com uma variável aleatória Poisson de parâmetro $\lambda = 1$ por m^2. Uma chapa é sorteada ao acaso para ser inspecionada, pergunta-se a probabilidade de:

A) Encontrarmos pelo menos 1 defeito.

B) No máximo 2 defeitos serem encontrados.

C) Encontrarmos de 2 a 4 defeitos.

D) Não mais de 1 defeito ser encontrado.

Exercício 6: Em momentos de pico, a chegada de aviões a um aeroporto se dá segundo o modelo Poisson com taxa de 1 por minuto. Determine a probabilidade de 3 chegadas em um minuto qualquer do horário de pico.

Outros modelos discretos

	Função de Probabilidade	Esperança	Variância
Geométrica(p)	$P(X = k) = p(1-p)^k$	${(1-p)}/{p}$	${(1-p)}/{p^2}$
Hipergeométrica(n, m, r)	$P(X = k) = {\binom{m}{k}\binom{n-m}{r-k}}/{\binom{n}{r}}$	${rm}/{n}$	$\frac{rm(n-m)(n-r)}{n^2(n-1)}$

- **Geométrica**: número de ensaios antes de ocorrer o primeiro sucesso (probabilidade de sucesso p).
- **Hipergeométrica**: número de objetos de um determinado tipo selecionados numa amostra de tamanho r de um conjunto de tamanho n com m objetos desse tipo.

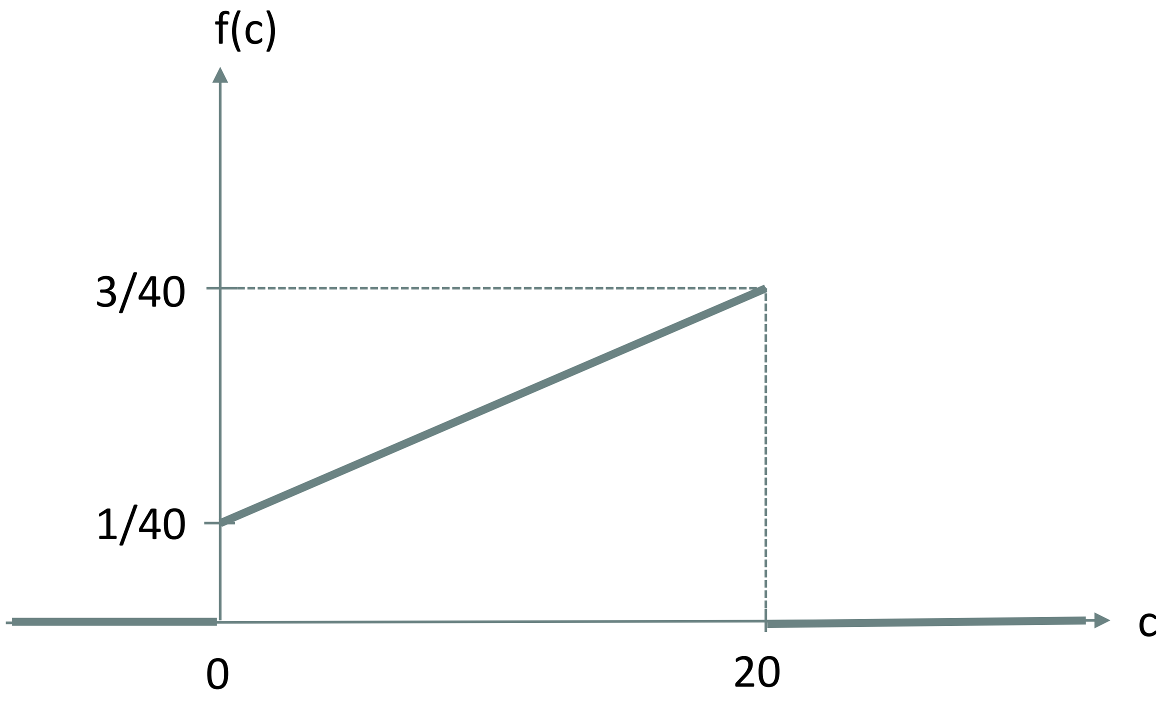
Função densidade de probabilidade

X é uma variável aleatória contínua se assume valores em intervalos dos número reais.

Exemplo: Arqueólogos estudaram uma certa região e estabeleceram um modelo teórico para a variável C, comprimento de fósseis da região (em cm). Suponha que C é uma **variável aleatória contínua** com a seguinte **função densidade de probabilidade**:

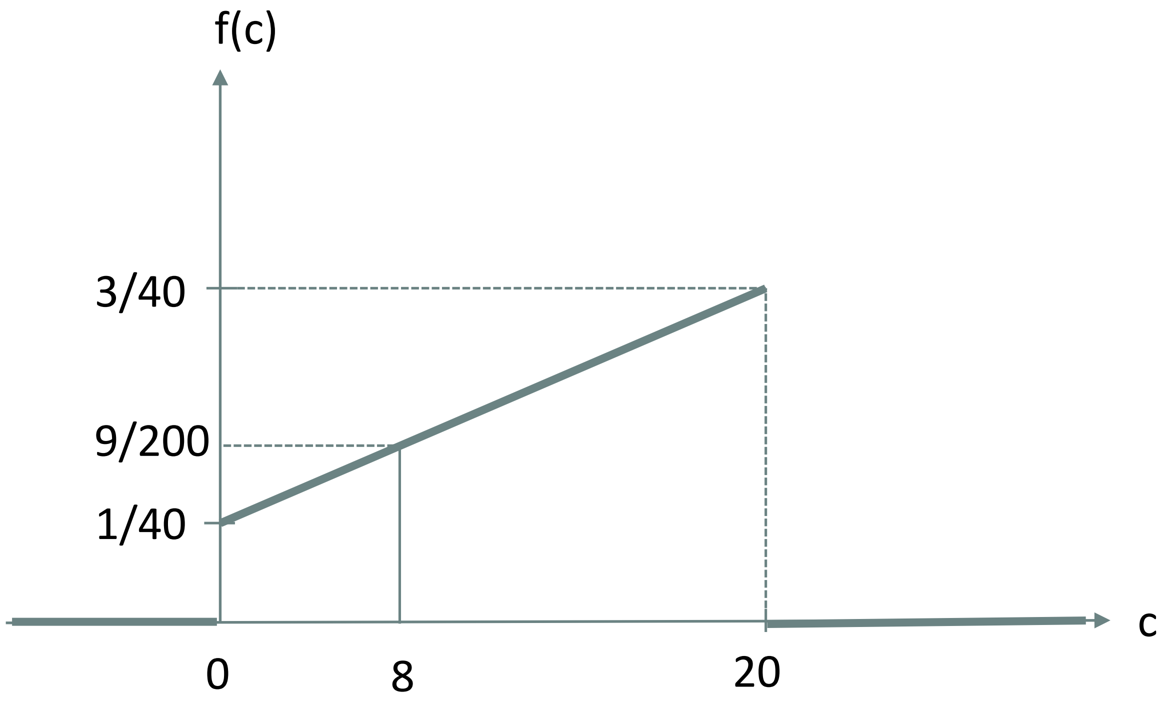
$$f(c) = \begin{cases} \frac{1}{40}\left(\frac{c}{10} + 1\right) \ se\ 0 \leq c \leq 20; \\ 0\ caso\ contrário \end{cases}$$

Função densidade de probabilidade



$$\acute{a}rea\ sob\ f(c) = \frac{\frac{1}{40} + \frac{3}{40}}{2} * 20 = 1$$

Função densidade de probabilidade



Probabilidade do comprimento ser menor do que 8 cm

$$P(C < 8) = \frac{\frac{1}{40} + \frac{9}{200}}{2} * 8 = \frac{7}{25}$$

Função densidade de probabilidade

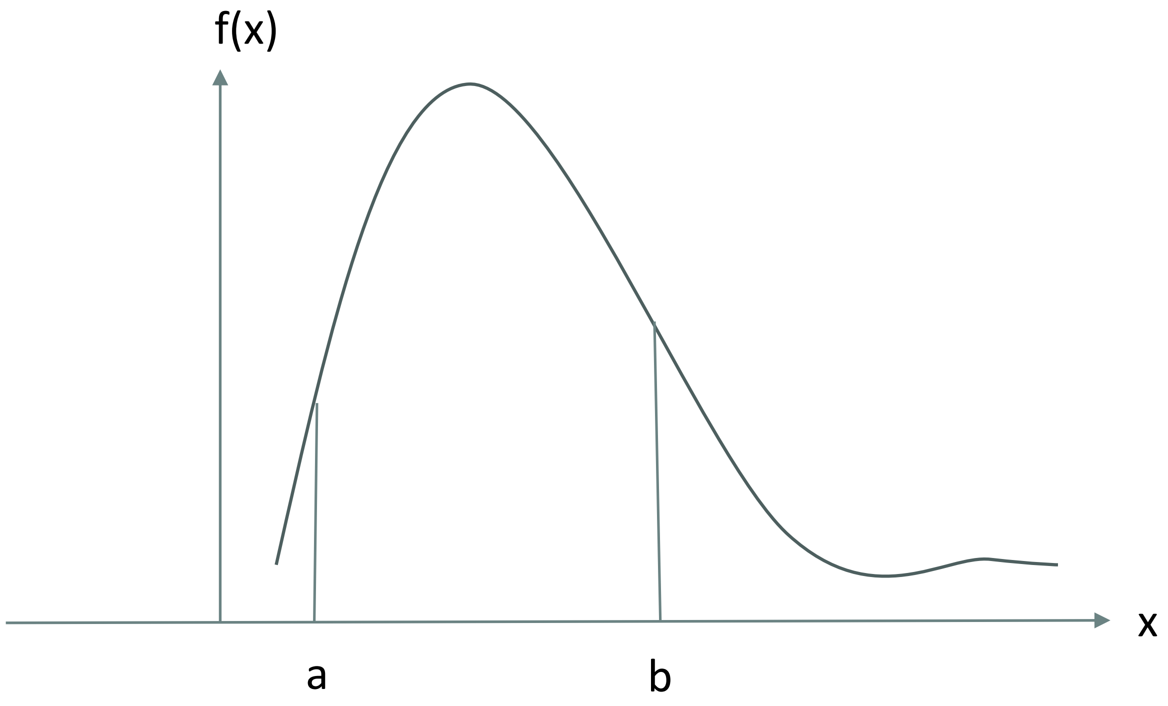
Dizemos que $f(x)$ é uma função densidade de probabilidade para uma variável aleatória contínua X, se satisfaz duas condições:

- $f(x) \geq 0$, para todo x
- A área definida por $f(x)$ é igual a 1

Observação: A segunda condição pode ser escrita como:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Função densidade de probabilidade acumulada



- Cálculo da probabilidade de $\mathrm{a} \leq X \leq b$:

Área abaixo do curva do ponto a ao ponto b, ou:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

- **Função densidade de probabilidade acumulada**:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Observação: $P(X = k) = 0$, para qualquer k, pois a área sob qualquer valor individual é zero

Esperança e Variância

A esperança de uma variável aleatória contínua:

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

A expressão acima é próxima à expressão para o caso discreto, apenas com a substituição da somatória pela integral e da probabilidade por $f(x)dx$.

A variância de uma variável aleatória contínua:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

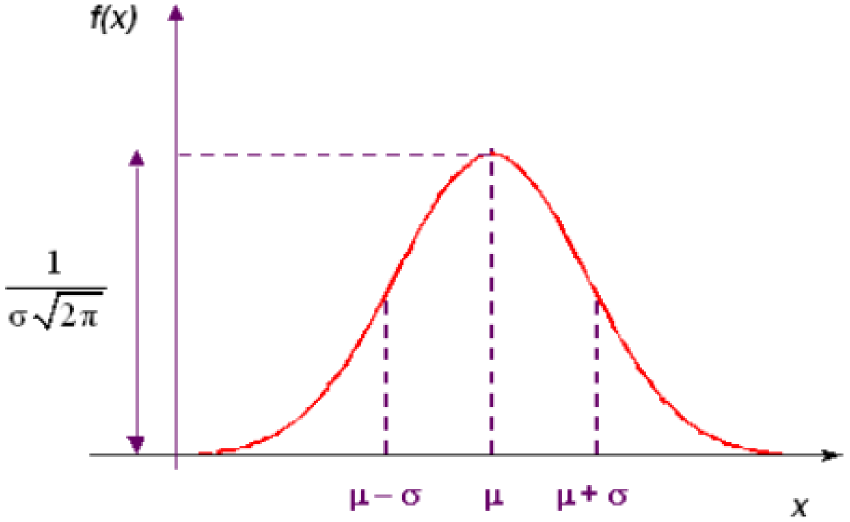
Modelo Normal

Dizemos que X tem distribuição Normal com parâmetros μ e σ^2, se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, -\infty < x < \infty$$

Dizemos que $X \sim N(\mu, \sigma^2)$.

- $E(X) = \mu$
- $Var(X) = \sigma^2$



Modelo Normal

Considere $X \sim N(\mu, \sigma^2)$ e $Z = \frac{X-\mu}{\sigma}$:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$
- $Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = \frac{1}{\sigma^2} * \sigma^2 = 1$

$Z \sim N(0,1): Normal\ Padrão$

Modelo Normal - Exemplos

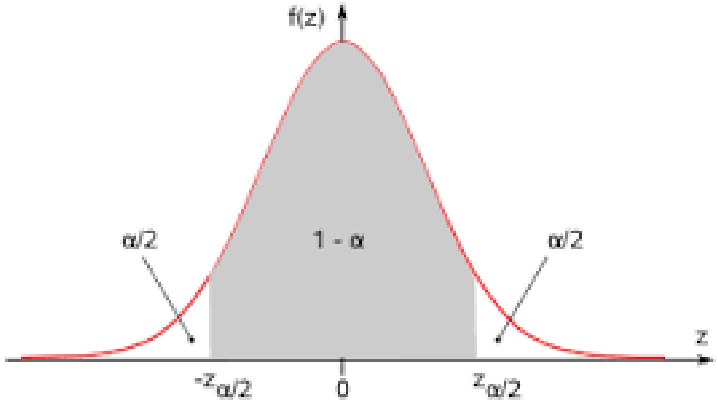
Seja $X \sim N(2{,}9)$, calcule a probabilidade:

A) $P(2 < X < 5) = \mathrm{P}\left(\frac{2-2}{\sqrt{9}} < Z < \frac{5-2}{\sqrt{9}}\right) = P(0 < Z < 1) = 0{,}3413$

B) $P(0 < X < 2) = P\left(\frac{0-2}{\sqrt{9}} < Z < \frac{2-2}{\sqrt{9}}\right) = P\left(-\frac{2}{3} < Z < 0\right) = P\left(0 < Z < \frac{2}{3}\right) = 0{,}2475$

C) $P(X < 3) = P\left(Z < \frac{3-2}{\sqrt{9}}\right) = P\left(Z < \frac{1}{3}\right) = 0{,}5 + 0{,}1293 = 0{,}6305$

D) $P(X > 4) = P\left(Z > \frac{4-2}{\sqrt{9}}\right) = P\left(Z > \frac{2}{3}\right) = 0{,}5 \ - 0{,}2486 = 0{,}2524$



Exercícios

Exercício 7: Seja $X \sim N(4,1)$, determine:

A) $\mathbb{P}[X \leq 4]$.

B) $\mathbb{P}[4 < X < 5]$.

C) $\mathbb{P}[2 \leq X < 5]$.

D) $\mathbb{P}[5 \leq X \leq 7]$.

E) $\mathbb{P}[X \leq 1]$.

F) $\mathbb{P}[0 \leq X \leq 2]$.

Exercícios

Exercício 8: Doentes, sofrendo de certa moléstia, são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal de média 15 e desvio-padrão 2 (em dias). Seja X o tempo de cura, temos que $X \sim N(15, 4)$.

A) Qual a probabilidade de um paciente demorar mais de 17 dias para se recuperar?

B) Qual a probabilidade de um paciente demorar menos de 20 dias?

Exercício 9: Os depósitos efetuados em um banco durante o mês de janeiro são distribuídos normalmente com média de $10.000,00 e desvio-padrão de $1.500,00. Um depósito é selecionado ao acaso no mês de janeiro. Encontre a probabilidade de que o depósito seja:

A) $10.000 ou menos.

B) pelo menos $10.000

C) Um valor entre $12.000 e $15.000

D) Maior do que $20.000

Aproximação Normal para o Modelo Binomial

Exemplo: Estudo do sindicato dos Bancários indica que cerca de 30% dos funcionários de banco têm problemas de estresse, provenientes das condições de trabalho. Numa amostra de 200 bancários, qual seria a probabilidade de pelo menos 50 com essa doença?

Admitindo que esses bancários foram sorteados ao acaso dentre todos os trabalhadores dessa categoria, podemos considerar que cada um deles tem a mesma probabilidade de ter ou não problema de estresse. Assim, X é o número total de bancários com problema dentre os 200 funcionários:

$$X \sim Bin(200; 0{,}3)$$

Queremos calcular:

$$P(X \geq 50) = \sum_{k=50}^{200} P(X = k) = \sum_{k=50}^{200} \binom{200}{k} 0{,}3^k 0{,}7^{200-k} \approx 0{,}9304$$

Aproximação Normal para o Modelo Binomial

Como a obtenção desse resultado será trabalhoso, podemos obter uma probabilidade aproximada através da distribuição Normal.

$$\left.\begin{aligned} &X \sim Bin(200; 0{,}3) \\ &E(X) = 200 * 0{,}3 = 60 \\ &Var(X) = 200 * 0{,}3 * 0{,}7 = 42 \end{aligned}\right\} \quad Y \sim N(\mu = 60; \sigma^2 = 42)$$

$$P(X \geq 50) \approx P(Y \geq 50) = P\left(Z \geq \frac{50-60}{\sqrt{42}}\right) = P(Z \geq -1{,}54) = 0{,}5 + 0{,}4386 = 0{,}9386$$

Aproximação Normal para o Modelo Binomial